

FACULTY OF ENGINEERING AND INFORMATION
TECHNOLOGY

Intelligent Early Warning System for Avian Influenza

Jie Zhang

A thesis submitted for the Degree of
Doctor of Philosophy



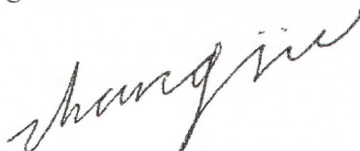
University of Technology, Sydney
April, 2011

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

A handwritten signature in cursive script, appearing to read 'Shangjie', is written above a horizontal line.

Acknowledgements

I would like to express my earnest thanks to my principal supervisor, Professor, Dr. Jie Lu, for her precious guidance and most generous help during the three and a half years of my doctoral research. Her comprehensive guidance has covered all aspects including the initial idea, the organizations of the papers and even the sentence structure and academic writing skills. Her comments are critical, accurate and challengeable. Her strict academic attitude and respectful personality has benefited my PhD study and will be a great treasure throughout my life. I also wish to express my sincere thanks to my co-supervisor, A/Professor, Dr. Guangquan Zhang for his knowledgeable suggestions and wise advice. Without their careful supervision and continuous encouragement, this research could not have been finished on time. Thanks to you all for your kind help.

I wish to thank Professor Xiaowei Yang from the School of Mathematical Science, South China University of Technology, for his precious help in comprehending the nature of the support vector machine method; finally, I can establish and implement the support vector regression algorithm. Also, I am grateful to Nature News reporter Dr. Declan Butler for his kind help in providing the important avian influenza outbreak data from 2003 to 2006.

I would like to thank Ms. Barbara Munday and Ms. Sue Felix for helping me correct the English presentation in my papers.

I am grateful to all members of the Decision Systems and e-Service Intelligent (DeSI) Lab for their careful participation in my presentation and valuable comments for my research.

I am grateful to the School of Software in the Faculty of Engineering and Information Technology at the University of Technology, Sydney. This study was fully supported by

an International Postgraduate Research Scholarship (IPRS) and UTS President’s Scholarship (UTSP) scholarship.

I would like also to thank my family members. Thanks to my mother and father for their conscious encouragement and generous support. Thanks to my daughter for her lovely eyes; she is everything to me.

Abstract

With the number of natural disasters has increased dramatically during the last decade, the early warning system (EWS) has become a necessary aid for all humankind in detecting incoming threats in good time, taking countermeasures beforehand and finally, mitigating the risks.

This research focuses on an intelligent epidemic EWS in the context of avian influenza. Computational intelligence (CI) techniques can provide cutting edge for an efficient and effective avian influenza EWS. The literature review reveals that the use of CI techniques in EWS is neither balanced nor systematic. This research proposes a conceptual framework and a technical framework as a guideline for integrating suitable CI techniques into an EWS from the aspects of structure, function and process. Following this guideline, we provide a hybrid knowledge-based prediction method which seamlessly connects case-based reasoning (CBR) and a fuzzy logic system to apply both implicit case knowledge and explicit expert knowledge in early warning prediction. In order to establish early warning in both a specific time and area, this research also puts forward two methods to address the issue. The first method is a seasonal auto-regressive based support vector regressive (SAR-SVR) time series prediction method, which applies SAR and Fast Fourier Transformation as the heuristic feature selection, and applies SVR to improve prediction accuracy. The second method employs one class classification (OCC) models by revising model combining policy and joining sub-classifiers OCC (JSC-OCC) methodology to realize the area risk mapping. Each method is followed by a validation with real world dataset.

Finally, an avian influenza intelligent early warning system (IEWs) prototype is implemented. The data used in the prototype system is real data collected from the

Internet, and thus the system could act as a validation means for this research. This prototype instantiates all the proposed approaches which can both estimate a risk level at a concrete location and map risk in a specific area in a specific time. The system realizes the consideration of involving suitable CI techniques in an EWS to form an IEWS with efficiency and effectiveness.

TABLE OF CONTENTS

CERTIFICATE OF AUTHORSHIP/ORIGINALITY i

Acknowledgements..... ii

Abstract..... iv

Table of Contents vi

List of Figures..... x

List of Tables xiii

CHAPTER 1 Introduction 1

1.1 Background 1

1.2 Research Objectives and Significance 2

1.3 Research Perspective and Methodology 7

1.3.1 Research Perspective..... 8

1.3.2 Research Methodology..... 9

1.4 Thesis Structure..... 10

1.5 Publications Related to This Thesis 12

CHAPTER 2 Literature Review..... 14

2.1 Early Warning Systems..... 14

2.1.1 Early Warning System Categories 14

2.1.2 Early Warning System Models 16

2.2 Avian Influenza Epidemic Early Warning System 18

2.2.1 Various Epidemic Early Warning Systems..... 18

2.2.2 Avian Influenza Epidemic Early Warning System	20
2.3 CI Methods for Early Warning System.....	25
2.3.1 CBR Method	26
2.3.2 Fuzzy Logic and Fuzzy Inference	27
2.3.3 Time Series Regression.....	30
2.3.4 One Class Data Description Model.....	33
CHAPTER 3 <i>Early Warning System Framework</i>	36
3.1 The Limitations of Existing Early Warning System Frameworks.....	36
3.2 Early Warning System Conceptual Framework Standards and Design.....	38
3.2.1 Socio-Technological Factors.....	38
3.2.2 Warning Lead-time	39
3.2.3 Feedback and Evaluation	39
3.3 A Conceptual Framework of Early Warning System	40
3.4 Development of the Intelligent Early Warning System Technical Framework.....	42
3.5 Design of an Knowledge Base for an Intelligent Early Warning System	44
3.6 Summary	47
CHAPTER 4 <i>A Hybrid Knowledge-based Prediction Method</i>	48
4.1 Avian Influenza Factors	48
4.2 Fuzzy Risk Level Identification.....	51
4.2.1 Fuzzy Risk Level.....	51
4.2.2 Identify a Fuzzy Risk Level	53
4.3 A Knowledge-Based Risk Prediction Method Using Fuzzy Rules and CBR ...	58
4.3.1 Motivation	59
4.3.2 A Knowledge-Based Risk Prediction Method Details	60
4.4 An Example	75
4.5 Validation	79
4.6 Summary	81

CHAPTER 5	<i>A Temporal SAR-SVR Prediction Method</i>	82
5.1	Box-Jenkins and Support Vector Regression Time Series Prediction Models (Preliminary)	83
5.1.1	Box-Jenkins Model	83
5.1.2	Support Vector Regression Model	83
5.2	Seasonal Auto-regressive Model Based Support Vector Regression Method..	86
5.3	A Case Study	92
5.3.1	Time Series Data Source of Avian Influenza	93
5.3.2	Data Pre-processing Method	93
5.3.3	Prediction Procedures and Results	94
5.4	Comparison with Other Models	101
5.5	Analysis and Discussions	111
5.6	Summary	115
CHAPTER 6	<i>One Class Classification Model for Risk Mapping</i>	117
6.1	OCC Methods and Feature Section	117
6.1.1	OCC Methods	118
6.1.2	Feature Selection Issue in OCC	121
6.2	A Combining OCC Method	122
6.3	The Combining OCC Method Experiments	125
6.4	Joint Sub-classifier OCC Method	130
6.4.1	The Motivation of JSC-OCC Method	130
6.4.2	Processes of the JSC-OCC Method	132
6.5	JSC-OCC method Experiment Results and Analysis	135
6.6	JSC-OCC method Performances Discussion	138
6.7	Summary	143
CHAPTER 7	<i>Intelligent Early Warning System Prototype</i>	145
7.1	Intelligent Early Warning System Structure	145

7.2 The Functions of an Intelligent Early Warning System 148

7.2.1 CBR Inference Function..... 148

7.2.2 SAR-SVR Prediction Function 152

7.2.3 OCC Risk Mapping Function..... 153

7.3 Data Source Descriptions..... 155

7.4 Summary 158

CHAPTER 8 Conclusions and Further Study..... 159

8.1 Conclusions..... 159

8.2 Further Study..... 161

References 164

Abbreviations..... 185

Index..... 189

LIST OF FIGURES

<i>Figure 1-1 Research perspective diagram.....</i>	<i>8</i>
<i>Figure 1-2 Thesis structure.....</i>	<i>11</i>
<i>Figure 2-1 Balancing risk by varying decision thresholds for evacuation orders (Little et al. 2007)</i>	<i>17</i>
<i>Figure 2-2 Diagrammatic summary of transmission routes of avian influenza viruses, showing major transmission routes (thick arrows) and minor transmission routes (thin arrows). (Peterson & Williams 2008).....</i>	<i>23</i>
<i>Figure 2-3 The case-based reasoning cycle (Aamodt & Plaza 1994)</i>	<i>27</i>
<i>Figure 2-4 The fuzzy inference system example</i>	<i>29</i>
<i>Figure 2-5 Rule 1 evaluation</i>	<i>29</i>
<i>Figure 2-6 Rule 2 evaluation</i>	<i>30</i>
<i>Figure 2-7 Rule combinations</i>	<i>30</i>
<i>Figure 3-1 The four elements of people-centred early warning systems (ISDR 2006, p2)</i>	<i>37</i>
<i>Figure 3-2 The general components of an integrated warning system (Mileti 1990, p2-4)</i>	<i>38</i>
<i>Figure 3-3 An EWS conceptual framework</i>	<i>41</i>
<i>Figure 3-4 The technical IEWS framework</i>	<i>43</i>
<i>Figure 3-5 The Knowledge base structure design</i>	<i>45</i>
<i>Figure 3-6 Relationship of case base and GIS database</i>	<i>46</i>
<i>Figure 3-7 Relationship among different data sets and different prediction methods</i>	<i>46</i>
<i>Figure 4-1 Factors associated with the outbreak</i>	<i>50</i>
<i>Figure 4-2 Risk level fuzzy function.....</i>	<i>52</i>
<i>Figure 4-3 The SEIR flow diagram.....</i>	<i>54</i>
<i>Figure 4-4 Simulation results</i>	<i>56</i>
<i>Figure 4-5 Poultry density rule pattern</i>	<i>62</i>
<i>Figure 4-6 Fuzzy membership functions of poultry density.....</i>	<i>63</i>

Figure 4-7 Fuzzy membership functions of distance to the nearest outbreak farm	64
Figure 4-8 A Poultry density rule set.....	65
Figure 4-9 A free range duck rule set.....	65
Figure 4-10 Harvest status membership functions	66
Figure 4-11 Free range poultry rule set.....	68
Figure 4-12 A wild bird migratory rule set	69
Figure 4-13 Membership function of distance to the nearest migratory place	69
Figure 4-14 Fuzzy membership functions of wild bird stay interval	70
Figure 4-15 Fuzzy membership functions of wild bird density.....	70
Figure 4-16 Comparison of two fuzzy numbers.....	75
Figure 4-17 Final risk level Fuzzy number results.....	78
Figure 5-1 The SVR-based method process.....	92
Figure 5-2 Avian influenza animal event weekly counts data	93
Figure 5-3 FFT and ACF results.....	95
Figure 5-4 ACF and PACF results (a) The autocorrelation function (ACF) and (b) partial autocorrelation function (PACF) of data series	96
Figure 5-5 The RMSE results of SAR-SVR	98
Figure 5-6 SAR-SVR prediction results with different parameters	100
Figure 5-7 Comparing the prediction results of the optimal models.....	108
Figure 5-8 Comparing the performances different models	110
Figure 5-9 Comparing the proposed SAR-SVR methods with SVR models.....	113
Figure 5-10 Comparing SAR-SVR with all time lags and SAR-SVR with selected time lags in set S.....	115
Figure 6-1 Combining OCC method process	124
Figure 6-2 One class SVDD vs. three separated sub-classes SVDD.....	132
Figure 6-3 The JSC-OCC method processes.....	134
Figure 6-4 JSC-OCC method on k-means clusters.....	141
Figure 6-5 JSC-OCC method on spectral clusters	142
Figure 7-1 IEWS prototype system structure.....	146
Figure 7-2 Case entry interface.....	149
Figure 7-3 Rules input interface.....	150

Figure 7-4 Inference interface 151

Figure 7-5 SAR-SVR prediction interface..... 153

Figure 7-6 JSC-OCC method risk mapping initial interface..... 154

Figure 7-7 JSC-OCC method risk mapping results interface..... 155

Figure 7-8 Major eight wild bird migratory routes..... 157

LIST OF TABLES

<i>Table 4-1 Fuzzy risk triangular fuzzy numbers and λ-cut.....</i>	<i>53</i>
<i>Table 4-2 Poultry infected fuzzy membership functions</i>	<i>57</i>
<i>Table 4-3 Wild bird infected fuzzy membership functions</i>	<i>57</i>
<i>Table 4-4 Poultry trade infected fuzzy membership functions.....</i>	<i>57</i>
<i>Table 4-5 Rules list</i>	<i>58</i>
<i>Table 4-6 Poultry density fuzzy numbers</i>	<i>63</i>
<i>Table 4-7 Fuzzy numbers of distance to the nearest outbreak farm.....</i>	<i>63</i>
<i>Table 4-8 A poultry density rule set.....</i>	<i>64</i>
<i>Table 4-9 Poultry density rule set cross table</i>	<i>64</i>
<i>Table 4-10 Harvest status fuzzy numbers</i>	<i>66</i>
<i>Table 4-11 A Free range duck rule set</i>	<i>67</i>
<i>Table 4-12 Free range duck rule set cross table</i>	<i>67</i>
<i>Table 4-13 Fuzzy numbers of distance to the nearest migratory place</i>	<i>69</i>
<i>Table 4-14 Fuzzy numbers of wild bird stay interval</i>	<i>70</i>
<i>Table 4-15 Wild bird density fuzzy number</i>	<i>70</i>
<i>Table 4-16 A Migratory birds rule set.....</i>	<i>71</i>
<i>Table 4-17 The cases examples.....</i>	<i>76</i>
<i>Table 4-18 Distances among different cases</i>	<i>76</i>
<i>Table 4-19 Rules preconditions and case parameters.....</i>	<i>77</i>
<i>Table 4-20 Fuzzy rule set one</i>	<i>77</i>
<i>Table 4-21 Part of fuzzy rule set three.....</i>	<i>77</i>
<i>Table 4-22 Results of applying fuzzy rules</i>	<i>77</i>
<i>Table 4-23 Distances to risk levels</i>	<i>78</i>
<i>Table 4-24 Validation results on Wisconsin breast cancer dataset.....</i>	<i>80</i>
<i>Table 5-1 Top 3 optimal performance of different SAR(p,P)s+SVR methods</i>	<i>99</i>
<i>Table 5-2 The detailed results of ARMA models</i>	<i>101</i>
<i>Table 5-3 The detailed results of top 5 SARMA models</i>	<i>102</i>

Table 5-4 The detailed optimal results of NN models..... 103

Table 5-5 The LS_SVR top 10 results 104

Table 5-6 The ϵ -SVR ($\epsilon=0.01$) top 10 results..... 105

Table 5-7 The ν -SVR($\nu=0.5$) top 10 results 105

Table 5-8 The SAR+LS-SVR top 10 results 106

Table 5-9 The SAR+ ϵ -SVR top 10 results 106

Table 5-10 The SAR+ ν -SVR top 10 results..... 107

Table 5-11 The summary of optimal results of different prediction methods 111

Table 6-1 OCC model experiment data-set description 125

Table 6-2 Experiment results with single OCC models 127

Table 6-3 Experiment results with combining 5 OCC models..... 128

Table 6-4 Experiment results with combining 6 OCC models..... 128

Table 6-5 Experiment results with combining 6 OCC models..... 129

Table 6-6 JSC-OCC experiment dataset..... 135

Table 6-7 Experiment with all features..... 137

Table 6-8 Experiment results with 12 features 137

Table 6-9 K-means: Three sub-groups 139

Table 6-10 Spectral clustering: Three sub-groups 139

Table 7-1 All the selected features..... 156

Table 7-2 Land cover name descriptions..... 158